Leading Edge
Essay
Cell

Green Genes—Comparative Genomics of the Green Branch of Life

John L. Bowman,[1,*] Sandra K. Floyd,[1] and Keiko Sakakibara[1]
[1]School of Biological Sciences, Monash University, Clayton Campus, Melbourne, Victoria 3800, Australia
*Correspondence: john.bowman@sci.monash.edu.au
DOI 10.1016/j.cell.2007.04.004

As more plant genome sequences become available, researchers are increasingly using comparative genomics to address some of the major questions in plant biology. Such questions include the evolution of photosynthesis and multicellularity, the developmental genetic changes responsible for alterations in body plan, and the origin of important plant innovations such as roots, leaves, and vascular tissue.

All plants are descended from a single eukaryotic ancestor that acquired a photosynthetic cyanobacterium as an endosymbiont (the ancestral plastid). The acquisition of a cyanobacterial endosymbiont was a momentous event in the evolution of life on Earth leading to a shift of most primary production from prokaryotic cyanobacteria to photosynthetic eukaryotes. Although the endosymbiosis of a cyanobacterium was a singular event in the history of life, plastids have also been transmitted horizontally to other eukaryotic lineages via secondary endosymbiotic events where unrelated eukaryotes acquired endosymbiotic plants. There are five or so eukaryotic lineages, one of which is plants (Keeling et al., 2005). Within the plants, three distinct groups have been identified (Figure 1): the glaucophytes (little-known freshwater algae), rhodophytes (red algae), and the green plants (which include green algae and land plants). The rhodophytes are primarily marine algae and include reef-building coralline algae; they provide a source of agar and form the basis of the billion-dollar nori industry in Japan. The green plants, by far the most diverse of the three groups, comprise two major clades: the chlorophytes (freshwater and marine algae) and the streptophytes (including the paraphyletic charophycean freshwater algae and the land plants). It was the land plants (embryophytes) that colonized and eventually dominated terrestrial landscapes and whose evolution allowed the subsequent colonization of land by the metazoans. Plastid genome sequences are available for species in all major lineages of plants, and nuclear genome sequences have been determined for a red alga, two chlorophytes, and three distinct lineages of land plants. Here, we highlight some of the major evolutionary transitions in the evolution of land plants and some key questions that are beginning to be addressed by comparing genome sequences from a diverse range of plant species.

The Algal Origins of the Photosynthetic Eukaryotes

Following the capture of a cyanobacterium by the plant ancestor, the evolution of the endosymbiont genome was characterized by wholesale transfer of genetic material to the host nuclear genome, resulting in a reduction in the endosymbiont genome and an enrichment of the host nuclear genome. Plastid genome sequences are available from all major plant lineages, and perhaps surprisingly, the plastid genomes from all lineages are similar in size and gene content. The plastids of the glaucophytes (called cyanelles) still retain a peptidoglycan cell wall characteristic of the ancestral cyanobacterial endosymbiont and yet have genomes similar to those of other plants. The continuing nuclear bombardment of plastid-derived DNA is thought to have contributed significantly to the genome content of plants, with as much as 18% of the genome of the model flowering plant (angiosperm) *Arabidopsis thaliana* thought to have been derived from the cyanobacterial endosymbiont (Martin et al., 2002). The contribution of cyanobacterial genes to algal nuclear genomes has not been analyzed in detail, but substantial differences might have contributed to differing genomic trajectories in the major plant lineages.

Table 1. Plant Genomes for Which Sequence Is Available (circa March 2007)

	Size (Mb)	# of Genes
Cyanidioschyzon merolae (*unicellular red alga*)	16.5	5331
Ostreococcus tauri (*unicellular green alga*)	12.56	8166
Chlamydomomas reinhardtii (*unicellular green alga*)	136	>15,000
Physcomitrella patens (*moss*)	487	>20,000
Selaginella moellendorfii (*lycophyte*)	85?	?
Oryza sativa (*rice*)	389	41,000
Populus trichocarpa (*popular tree*)	485	45,000
Arabidopsis thaliana (*flowering plant*)	140	27,500

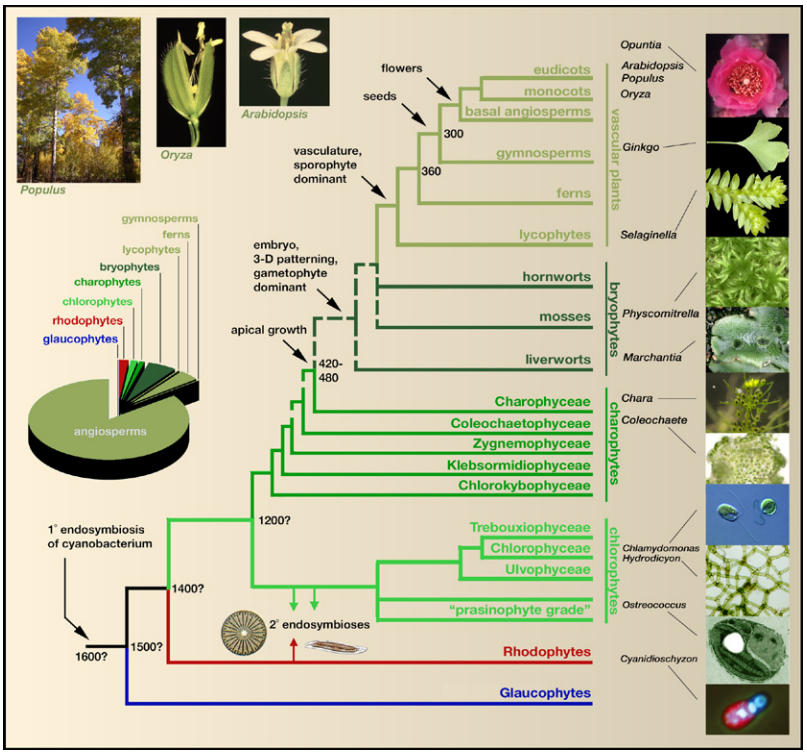

Figure 1. Phylogenetic Relationships among Plants

Depicted are relationships among the three lineages of plants: glaucophytes (freshwater algae; blue), rhodophytes (red algae; red), and the green plants (chlorophytes, charophytes, and land plants; green). Estimated dates for some nodes are listed in millions of years before present. The primary endosymbiotic event is estimated to have occurred at least 1.6 billion years ago. A deep split within the green lineage created the chlorophyte clade and the charophyte plus land plant clade. Note that both the charophytes and the bryophytes are grades and are not monophyletic. Major events in the evolution of land plants are demarcated with arrows. Species for which complete nuclear genome sequences are available are listed in color (photographs at right; the three angiosperm species are pictured upper left). Species positioned in large phylogenetic gaps where genome sequences would be informative (black) include the following: the basal lineage of land plants, the liverworts, charophycean algal lineages (*Chara*, *Coleochaete*) that are sisters to land plants, and the gymnosperms, which are the sister group to flowering plants (angiosperms). Also included is a multicellular chlorophytic green alga. Secondary endosymbiotic events have occurred within both the red algae (e.g., diatoms, pictured) and green plants. Pie chart shows the relative species richness of the major clades. The vast majority of species within the Plantae are angiosperms (250,000 species), with other groups having substantially fewer described species (numbers approximated): glaucophytes 13; rhodophytes 5,920; chlorophytes 3,720, charophytes 3,400; bryophytes 17,000 (liverworts 7,000, mosses 10,000, hornworts 100); lycophytes 1,225; ferns 12,000; gymnosperms 800.
Photos from top: *Opuntia basilaris*, *Ginkgo biloba*, *Selaginella kraussiana*, *Physcomitrella patens*, *Marchantia polymorpha*, *Chara* sp., *Coleochaete* sp., *Chlamydomonas reinhardtii*, *Hydrodictyon sp.*, *Ostreococcus tauri*, *Cyanidioschyzon merolae*. Photos courtesy of Gayle Dupper, Institute of Forest Genetics, Placerville, CA, USA (poplar), Charles Delwiche, University of Maryland (*Chara*), James Umen, Salk Institute (*Chlamydomonas*), Hervé Moreau, Université Pierre et Marie Curie-Paris (*Ostreococcus*), and Tsuneyoshi Kuroiwa, Rikkyo University (*Cyanidioschyzon*).

There are three plant species for which almost complete genome information for nucleus, chloroplast, and mitochondrion is available: the red alga *Cyanidioshyzon merolae*, the marine prasinophycean green alga *Ostreococcus tauri*, and the chlorophycean green alga *Chlamydomonas reinhardtii* (Table 1). These unicellular algae are ideal models for cell biology because the cells are monoplastidic, with *C. merolae* and *O. tauri* cells also containing only a single mitochondrion and Golgi body, the division of which can be synchronized. *C. merolae* lives in acidic hot springs but can be grown in culture. Its genome is compact with most genes lacking introns (Matsuzaki et al., 2004). The phytoplankton *O. tauri* is a picoeukaryote comprising cells that are about the size of prokaryotes (about 1 μm in diameter). Its genome is similarly compact with an average spacing of only 197 basepairs between genes (Derelle et al., 2006). One remarkable feature of the *O. tauri* genome is its extreme heterogeneity with 2 chromosomes differing from the other 18 in GC content and transposable element distribution, suggesting horizontal acquisition of at least one of its chromosomes. In contrast, the *C. reinhardtii* genome is larger and contains more genes. A comparative genomics study using the *Chlamydomonas*, *Arabidopsis*, and human genomes facilitated the identification of genes involved in flagellar development and function in both *Chlamydomonas* and humans, including genes involved in human disease (angiosperms lack the flagellated sperm found in many other organisms) (Li et al., 2004). *C. reinhardtii* is a sophisticated model for investigating photosynthesis and fundamental cell biology with tools available for transformation of all three genomes (nuclear, chloroplast, and mitochondrial) and for both forward and reverse genetics (reviewed in Grossman et al., 2007). Additional red and green algal genome sequences, such as the sequences of two other *Ostreococcus* genomes (US Department of Energy Joint Genome Institute, www.jgi.doe.gov), are required to assess whether these characteristics are unique or more general for these taxa.

By producing oxygen as a waste product, the evolution of photosynthetic cyanobacteria 3.5 billion years ago dramatically altered the Earth's ecosystem. Following the primary endosymbiotic event that defines plants, this eukaryotic lineage evolved to become the dominant primary producer in both aquatic and terrestrial habitats. Comparisons among algal genome sequences can provide information to elucidate characteristics of the ancestral photosynthetic eukaryotes. For example, the *O. tauri*

genome includes genes potentially involved in C4 photosynthesis, which enhances photosynthetic capabilities under low CO_2 conditions and has evolved repeatedly in several angiosperm lineages. C4 photosynthesis in *O. tauri* would confer a significant advantage under specific environmental conditions, suggesting that this capability may have been present at an early stage of green algal evolution (Derelle et al., 2006). Additionally, genomic comparisons between *C. merolae* and the green algae will provide insight into both shared and specific genetic characters in the two algal lineages.

Becoming Multicellular

The emergence of multicellular organisms from unicellular ancestors occurred repeatedly in the evolution of eukaryotes, most notably in the metazoan and land plant lineages. The origin of multicellularity is thus one of the key questions in the evolution of life on Earth. Comparative genomics suggests that a combination of co-opting existing genes for new functions and the evolution of new proteins from novel combinations of pre-existing protein domains contributed to the emergence of multicellularity in metazoans (Ruiz-Trillo et al., 2007). Multicellularity has evolved numerous times within the red and green algae. Are similar or distinct genetic programs recruited to pattern multicellular algal taxa? Is multicellularity in land plants fundamentally different or similar to that of their algal relatives? Comparative genomics of multicellular organisms from distinct lineages should shed light on these questions. However, given that only unicellular algae have thus far had their genomes sequenced, sequencing of the genomes of multicellular red and green algae will be required to address this issue.

Conquering the Land

The origin of land plants from aquatic ancestors marks a major evolutionary transition in the history of green plants. Land plants inherited many biochemical, ultrastructural, and physiological characters from their algal ancestors. Comparison of mitochondrial genomes of the charophycean alga *Chara*, the liverwort *Marchantia polymorpha*, and other land plants provides some of the strongest evidence for the sister relationship of *Chara* to land plants and of liverworts to other land plants (Turmel et al., 2003). Unlike the development of their closely related multicellular charophycean algal relatives, land plants exhibit growth from an apical meristem that produces a three-dimensional body that becomes patterned to produce distinct tissues. One of the key questions is how programs for development and growth were changed to allow the production and patterning of tissues. The ability to compare the ancestral land plant genome with that of algal relatives would facilitate the identification of the genetic bases for the key innovations that allowed green plants to evolve from aquatic ancestors and adapt to life on land. Such key innovations include the perception of environmental cues (light and gravity), the origin of extracellular matrices (sporopollenin, lignin, and pectic acid), establishment of intercellular communication networks (plasmodesmata, plant hormones, receptors, and their ligands), and diversification of gene regulatory networks promoting cell differentiation. Because of the enormous evolutionary divergence between chlorophytes and streptophytes, sequencing of the genome of a charophycean alga (such as *Chara*) will be required to assess the origins of genetic mechanisms in land plants.

The closest relatives of land plants, the charophycean algae, have a haplontic life cycle in which the zygote is the only diploid cell. All land plants have a life cycle that includes an alternation of generations involving a haploid phase (gametophyte) in which gametes are produced and a diploid phase (sporophyte) that produces spores. Thus, multicellularization of the zygote evolved early during land plant evolution. Was the initial elaboration of the zygote to produce a multicellular diploid sporophyte due to a co-option of already existing developmental pathways of gametophyte development or was it due to the origin of de novo developmental genes and networks? The earliest land plants most likely had a haploid-dominant life cycle, with an ephemeral-dependent sporophyte, and this has been retained in the extant bryophytes (mosses, hornworts, and liverworts). Flowering plant models such as *Arabidopsis*, rice, and poplar all represent diploid-dominant plants in which the sporophyte is long lived and complex and the gametophyte is diminutive and ephemeral. One of the major questions in plant evolution concerns the evolution of the sporophyte developmental program, which was modified through time so that sporophytes became larger and acquired the ability to branch, develop conducting tissues, and produce roots, leaves, seeds, and flowers.

Another key question concerns the relationship between radial and bilateral or dorsiventral development in land plants (Friedman et al., 2004). The most familiar instance of dorsiventral development is that of leaves. There has been a great deal of research interest in understanding the genetics of polarity and growth of leaves in flowering plants. Organs referred to as leaves occur in all extant vascular plants, but in at least three cases these leaves evolved independently, in lycophytes, ferns, and seed plants. The earliest vascular plants (known only from fossils) lacked laminar, lateral, vascularized appendages. Thus, in vascular plants, organs with dorsiventral polarity evolved in the sporophyte generation that had radial patterning mechanisms. Variability in growth form also exists in gametophytes: some have radial organization (mosses, whisk ferns) and others have dorsiventral or thalloid organization (liverworts, hornworts, ferns). Evidence from the earliest land plant fossils suggests that the earliest land plants may have been liverworts or liverwort-like plants with a thalloid gametophyte. If this is true, then a transition from dorsiventral to radial gametophyte development must have occurred within land plants, and more than once.

Liverworts represent the sister group to all other extant land plants. The best hope of assessing the nature of the land plant ancestral genome for comparison with algal genomes will require comparison of a liverwort genome with that of other land plant genomes. For example, the genome of the thalloid liverwort *M. polymorpha* will provide the basis for comparing the developmental genetics of plants with dorsiventral development and those with radial development. Efforts toward obtaining the nuclear genome sequence of *M. polymorpha* include the construction of BAC libraries (Green Plant BAC project), end-sequences of both BAC and EST libraries (T. Kohchi, personal communication), and submission of a pilot proposal for whole-genome shotgun sequencing (JGI).

Although genome sequences from liverworts and charophycean algae are not yet available, the nuclear genome of the moss, *Physcomitrella patens*, has recently been sequenced along with a large number of cDNA clones derived from various developmental stages, including leafy shoots of gametophytes and sporophytes (M. Hasebe, personal communication). Initial analyses of cDNA sequences suggest that mosses and angiosperms have largely the same types of gene families, including most of the gene families implicated in developmental patterning in angiosperms. This suggests a cooption of existing genes rather than the evolution of new genes in the transition from a gametophyte-dominant life cycle to a sporophyte-dominant one (Nishiyama et al., 2003; Floyd and Bowman, 2007). However, the gene families have markedly diversified in the angiosperms relative to mosses (Floyd and Bowman, 2007). Functional analyses using homologous recombination knockout technology in *P. patens* will be required to clarify the questions of whether the same genetic networks function in both haploid and diploid generations of land plants, whether the radial shoots of the moss gametophyte and the vascular plant sporophyte are regulated by similar developmental programs, and whether body plans in the different generations require different developmental programs.

Becoming Large, the Evolution of Vasculature

Another plant genome that has been sequenced and awaits assembly and annotation is that of the lycophyte *Selaginella moellendorffii*. There are two major lineages of extant vascular plants, the lycophytes (spike mosses, club mosses, quillworts) and the euphyllophytes (ferns, horsetails, seed plants), representing an ancient divergence of a vascular plant ancestor. These lineages separated prior to the evolution of many features we commonly associate with plants. Leaves, roots, and complex vascular architectures have evolved independently within both lineages from a morphologically simpler common ancestor. Despite millions of years of evolution, lycophytes have also retained many developmental features thought to be ancestral or primitive for vascular plants. These include an apical meristem with one or a few apical initial cells, apical dichotomous branching, and a protostelic vasculature with xylem surrounded by phloem. The differences between lycophytes and euphyllophytes highlight some of the major questions in vascular plant evolution. How were the complex shoot apical meristems of seed plants derived from simpler ancestral meristems? Are both simple and complex meristems regulated by the same gene regulatory networks? How might these networks have changed as the simpler ancestral meristems evolved? Are the independently acquired leaves and roots of these organisms patterned by the same or different genetic programs? Are the vascular tissues analogous or homologous to the conducting tissues in mosses?

In the case of leaves some insight has already been gained from genomic data in addition to using candidate gene approaches (Floyd and Bowman, 2006, Harrison et al., 2005). Although the *S. moellendorffii* genome has not been assembled yet, searches of the unassembled sequences have identified many developmental gene families that are shared with flowering plants and some that are not (reviewed in Floyd and Bowman, 2007). Two gene families important for leaf development in flowering plants, Class III HD-Zip and KANADI, are both present in the genome of *S. moellendorffii*. However, the subclade of Class III HD-Zip genes involved in leaf polarity in flowering plants has no ortholog in *S. moellendorffii*. Likewise, the YABBY gene family, important for abaxial identity and laminar outgrowth in flowering plants, has not been found in the *S. moellendorffii* genomic sequence. With the completion of the assembly and annotation of the *S. moellendorffii* genome the full assessment of many gene families will be possible and we can begin to address long-standing questions in vascular plant evolution with a new set of genetic tools. Techniques for genetic transformation in *S. moellendorffii* enabling transgenic approaches for studying gene expression may also be possible.

Genome Duplications and Morphological Innovations in Flowering Plants

One surprising discovery from the genome sequences of the model plants *A. thaliana* and rice (*Oryza sativa*) is evidence for repeated whole-genome duplications, despite the diploid nature of the two species. Flowering plants offer an attractive system to study the consequences of whole-genome duplications due to their propensity for polyploidization. Flowering plants have likely undergone multiple rounds of polyploidization in the past 150–200 million years. In contrast, in mammals polyploidization has been suppressed over the same timeframe due to the presence of the X-Y sex chromosome system. There are three key questions: (1) How does the process of diploidization occur? (2) Do whole-genome duplications correlate with speciation events? and (3) Do whole-genome duplications correspond to an explosive evolution of morphology by providing the raw material of entire genetic pathways for selection, as suggested by Ohno (1970)? The

availability of multiple whole-genome duplications of varying antiquity in flowering plants facilitates the formulation of some hypotheses to address these questions.

Several groups have tried to date the three whole-genome duplications (called 1R, 2R, and 3R) that have left their imprint in the organization of the *Arabidopsis* genome (reviewed in De Bodt et al., 2005). Although there is much uncertainty in the precise timings, the earliest event could correspond with the origin of extant flowering plants, and the second event with the origin of the eudicots (the most species-rich clade of flowering plants). Likewise, a whole-genome duplication event dating to the base of the monocots, the second large clade of angiosperms, is evident in the genome of rice. That the whole-genome duplications correlate in time with origins or major radiations of flowering plants is consistent with Ohno's idea that such events facilitate major leaps in morphological evolution. More precise dating of the whole-genome duplication events relative to the adaptive radiations of angiosperms will require the analysis of other phylogenetically informative angiosperm genomes such as that of the basal eudicot *Aquilegia* (columbine; in progress at JGI).

Based on comparison of the genome sequences of *Arabidopsis* and of the poplar tree (*Populus*), it has been estimated that their common ancestor possessed only 12,000–14,000 genes. This suggests differential retention of genes in the different lineages, possibly related to their different life histories, that is, ephemeral annual versus long-lived and sometimes vegetatively propagated perennial (Tuskan et al., 2006; Maere et al., 2005). Analysis of the *Arabidopsis* genome has also revealed the preferential retention of specific classes of genes—such as those encoding transcription factors, signaling molecules, and secondary metabolism enzymes—following whole-genome duplications (Blanc and Wolfe, 2004). Such preferential retention could also be evidence of a requirement to maintain an appropriate stoichiometry in protein complexes and a selection for increased diversity of secondary metabolites involved in defense. The preferential retention of transcription factors is also consistent with the higher percentage of transcription factors in multicellular plants (12%–15%) relative to unicellular plants (2%–4%). The increase in genes encoding signaling molecules and transcription factors is also consistent with the idea of neofunctionalization following gene duplication contributing to the evolution of morphological complexity. The evolutionary process of diploidization, whereby a polyploid decays to become a diploid, is enigmatic. However, the study of recent polyploids suggests that massive gene loss accompanied by structural evolution of chromosomes and epigenetic reprogramming of retained genes may influence chromosome pairing and thus contribute to diploidization.

The evolution of the flower and the carpel were key innovations that allowed angiosperms to engage in specialized animal pollination systems and seed dispersal mechanisms. Did the evolution of these innovations require the evolution of new genes, or were already existing genetic programs co-opted to new roles? Similar to the situation described earlier in the evolution of vascular plants from a bryophyte-like ancestor, orthologs of flower patterning genes are present in gymnosperms (that is, nonflowering seed plants such as conifers). For example, B and C class MADS box genes that pattern the reproductive organs of angiosperm flowers are also expressed in the reproductive organs of gymnosperms. However, whereas gymnosperms appear to have single copies of a C class gene, angiosperms harbor multiple copies derived from gene/genome duplications within the angiosperm lineage. At least two C class genes are present in all angiosperms, suggesting that neofunctionalization could have contributed to one gene being specialized for carpel development (an angiosperm-specific structure) and the other gene specialized for ovule and integument development (shared by both angiosperms and gymnosperms). In contrast, secondary growth (the production of wood) has secondarily evolved multiple times within the angiosperms from herbacious ancestors, suggesting that most if not all angiosperm species still possess the ancestral genetic programs. This implies that secondary growth may involve co-option of pre-existing genetic programs via changes in gene regulation mediated, for example, by modifications to chromatin. Thus, it is likely that both the evolution of new genes, via gene duplication events, and the co-option of existing genetic programs contributed to the evolution of morphological diversity within the angiosperms, and that the ample genetic material provided by whole-genome duplications has played a major role in the rise of the angiosperms as the dominant land plant vegetation on the planet today.

Conclusion

Genomes from plants representing different phylogenetic lineages, levels of organization, and body plan will soon be available for comparative genomic analyses and for functional analysis of development using reverse genetics and transgenic techniques. As researchers begin to mine the rich source of data from *Cyanidioshyzon*, *Ostreococcus*, *Chlamydomonas*, *Physcomitrella*, and *Selaginella* to compare with *Arabidopsis*, *Oryza*, and *Populus*, we look ahead to the addition of still more plant genomes such as those of *Marchantia* and *Chara* to bridge some of the vast evolutionary gaps that remain. We are on the verge of a new and exciting era of comparative genomics for the major lineage of photosynthetic organisms, and the future looks very green.

ACKNOWLEDGMENTS

Authors are supported by the US National Science Foundation and the Australian Research Council.

REFERENCES

Blanc, G., and Wolfe, K.H. (2004). Plant Cell *16*, 1679–1691.

De Bodt, S., Maere, S., and Van de Peer, Y. (2005). Trends Ecol. Evol. *20*, 591–597.

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., et al. (2006). Proc. Natl. Acad. Sci. USA *103*, 11647–11652.

Floyd, S.K., and Bowman, J.L. (2006). Curr. Biol. *16*, 1911–1917.

Floyd, S.K., and Bowman, J.L. (2007). Int. J. Plant Sci. *168*, 1–35.

Friedman, W.E., Moore, R.C., and Purugganan, M.D. (2004). Am. J. Bot. *91*, 1726–1741.

Grossman, A.R., Croft, M., Gladyshev, V.N., Merchant, S.S., Posewitz, M.C., Prochnik, S., and Spalding, M.H. (2007). Curr. Opin. Plant Biol. *10*, 190–198.

Harrison, C.J., Corley, S.B., Moylan, E.C., Alexander, D.L., Scotland, R.W., and Langdale, J.A. (2005). Nature *434*, 509–514.

Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., and Gray, M.W. (2005). Trends Ecol. Evol. *20*, 670–676.

Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y., Teslovich, T.M., May-Simera, H., Li, H., Blacque, O.E., Li, L., Leitch, C.C., et al. (2004). Cell *117*, 541–552.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. (2005). Proc. Natl. Acad. Sci. USA *102*, 5454–5459.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Proc. Natl. Acad. Sci. USA *99*, 12246–12251.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004). Nature *428*, 653–657.

Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., et al. (2003). Proc. Natl. Acad. Sci. USA *100*, 8007–8012.

Ohno, S. (1970). *Evolution by Gene Duplication* (George Allen and Unwin, London).

Ruiz-Trillo, I., Burger, G., Holland, P.W.H., King, N., Lang, B.F., Roger, A.J., and Gray, M.W. (2007). Trends Genet. *23*, 113–118.

Turmel, M., Otis, C., and Lemieux, C. (2003). Plant Cell *15*, 1888–1903.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., et al. (2006). Science *313*, 1596–1604.